Bayesian estimation, classification, and denoising based on alpha-stable statistical modeling

Panagiotis Tsakalides
FORTH-ICS &
Computer Science Department
University of Crete
Heraklion, Greece

www.ics.forth.gr/~tsakalid
tsakalid@ics.forth.gr

Before Starting

- I am not a physicist
- I am not an astronomer
- Closest ever been:

JPL Open House, Pasadena, CA 3-5-2008

Presentation Outline

- Heavy-tailed signals and non-Gaussian modeling
- Multiscale methods for SAR image processing
- Novel Bayesian processors for estimation and denoising
- Real images results and conclusions

Symmetric Alpha-Stable (SaS) Processes:

A (fairly…) New Statistical Signal Processing Framework

Quotation

"The tyranny of the normal distribution is that we run the world … by attributing average levels of competence to the whole population.

What really matters is what we do with the tails of the distribution rather that the middle."

R. X. Cringely
Accidental Empires, 1992

It can also be said about least-squares in signal processing.

The Symmetric Alpha-Stable (SaS) Model

SaS Characteristic Function:

$$\phi(\omega) = e^{j\delta\omega - \gamma|\omega|^{\alpha}}$$

α: characteristic exponent, 0<α<=2 *(determines thickness of the distribution tails, α=2: Gaussian, α=1: Cauchy)*

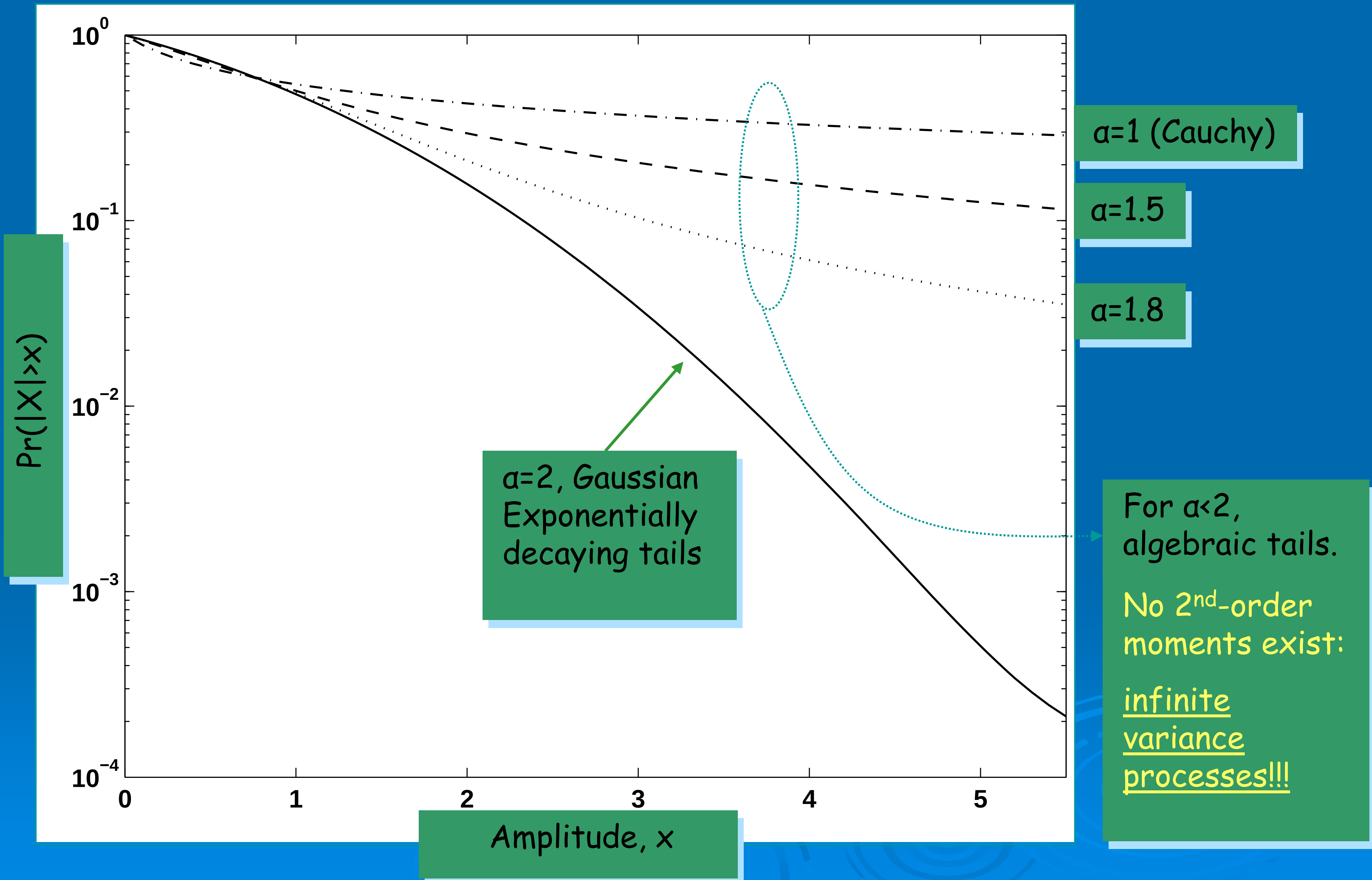
δ: location parameter *(determines the pdf's point of symmetry)*

γ: dispersion parameter, γ>0 *(determines the spread of the distribution around its location parameter)*

for Gaussian → γ = 2 x variance

for Cauchy → γ behaves like variance

SaS Probability Functions



Properties of SaS Laws

- Naturally arise as limiting processes via the Generalized Central Limit Theorem.
- Possess the stability property: The shape of a SaS r.v. is preserved up to a scale and shift under addition.
- Contain Gaussian (a=2) and Cauchy (a=1) distributions as members.
- Have heavier tails than the Gaussian: Their tail probabilities are asymptotically power laws → More likely to take values far away from the median ("Noah effect"):

$$P(X > x) \sim c_\alpha x^{-\alpha} \quad \text{as} \quad x \to \infty$$

Properties of SaS Laws

- Have finite *p*-order moments only for p<a:

$$Ex^p \prec \infty \quad \text{for} \quad p \prec \alpha$$

- Do not have finite second-order moments or variances:

$$Ex^2 = \infty$$

- Are self-similar processes: Exhibit long-range dependence or long memory ("Joseph effect").

Key Question!

- Since the variance if associated with the concept of power, are infinite variance distributions inappropriate for signal modeling and processing??

- No!! Variance is only one measure of spread! What really matters is an accurate description of the shape of the distribution. Particularly true when outliers appear in the data.

- Note that bounded data are routinely modeled by the Gaussian distribution, which has infinite support.

Fractional Lower-Order Moments

- Covariation of two alpha-stable r.v.'s:

$$[X,Y]_\alpha = \frac{E\left[XY^{\langle p-1\rangle}\right]}{E[|Y|^p]}\gamma_Y \quad \text{for } 1 \le p \prec \alpha,\ Y^{\langle\beta\rangle} = |Y|^{\beta-1}Y^*$$

- Fractional Lower-Order Moments:

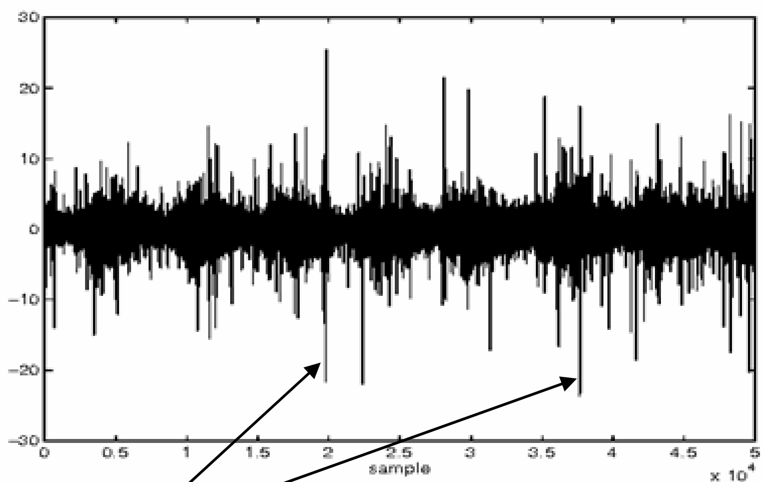
$$\left\langle X, Y \right\rangle_{p_1, p_2} = E\left[X^{\langle p_1\rangle} Y^{*\langle p_2\rangle}\right], \quad p_1 + p_2 \prec \alpha$$

- Estimation Algorithms based on *L_p-Norm Minimization,* where *1<= p < α <2.*

Real Data Modeling

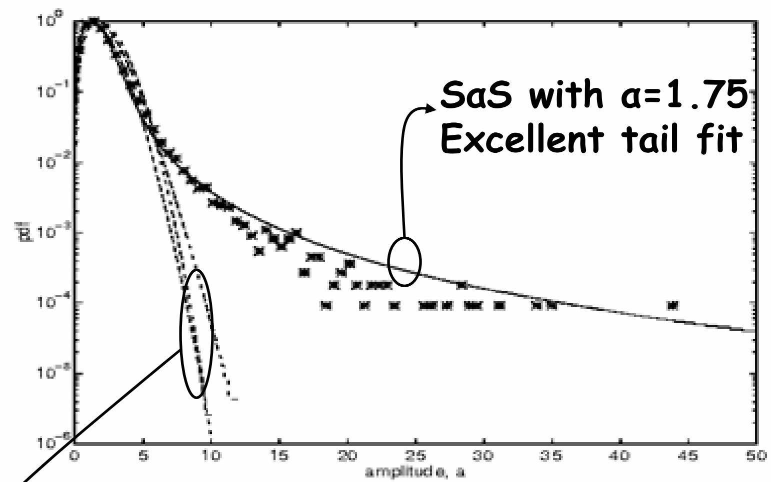
<u>Real sea clutter @ nominal sea condition:</u>

- sea state 3
- X-band radar
- 8o look-down angle
- spatial resolution of 1.52 m (5 ft)
- sampled at 40 Hz



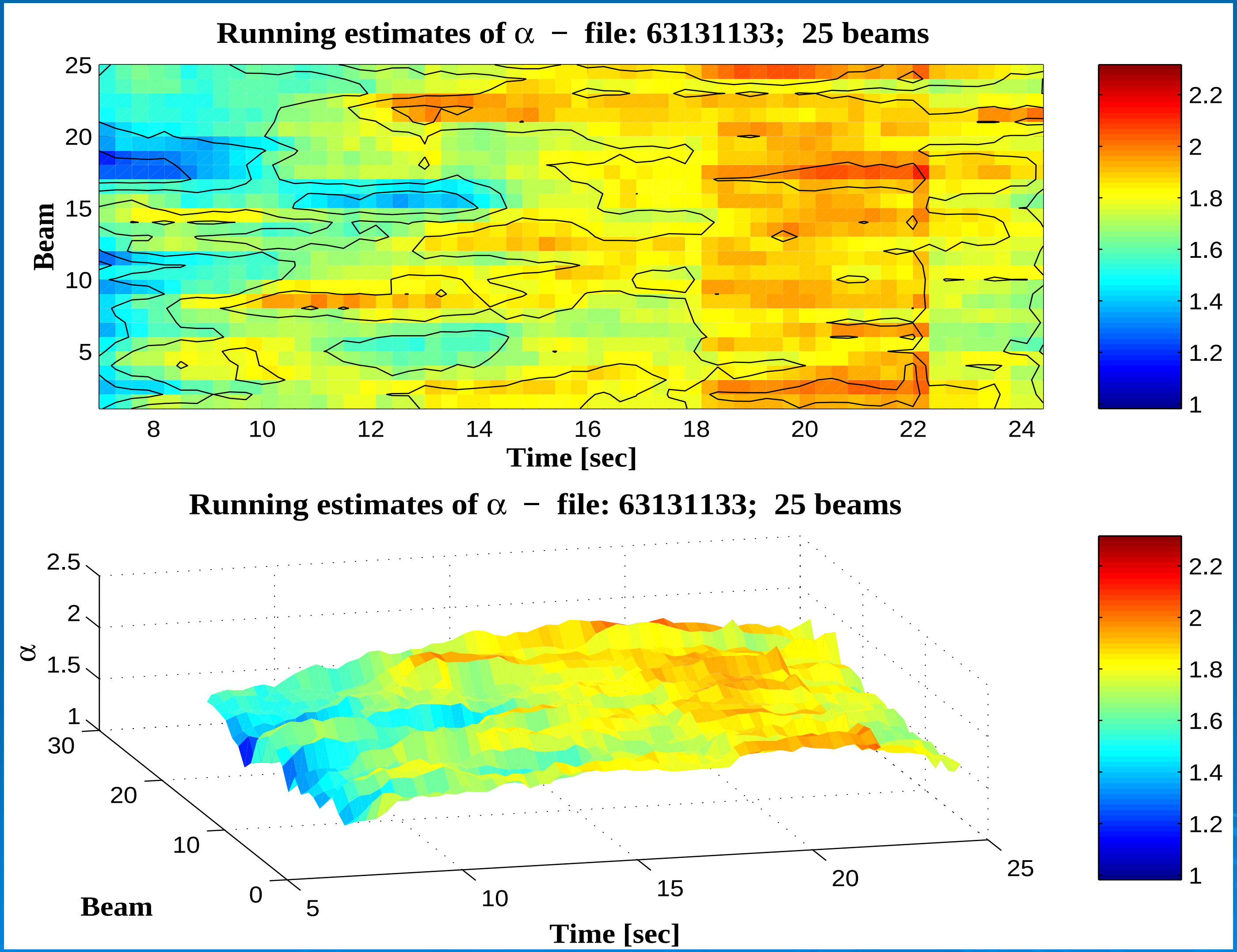
The impulsive nature of the clutter data is obvious.

<u>Clutter probability density modeling</u>



Exponential densities

Real Data Modeling



SaS Applications

- Economic Time Series (Mandelbrot 60's, McCulloch 90's)
- Statistics (Zolotarev, Cambanis, Taqqu, Koutrouvelis, 70's-90's)
- Modeling of Signals and Noise:
  - Radar clutter – The Cauchy Beamformer, the ROC-MUSIC Algorithm (Tsakalides and Nikias, 1995)
  - Underwater Noise - The alpha-matched filter (Tsakalides, 1997)
  - Communications Applications
    - Telephone line noise (Stuck and Kleiner, 1974)
    - Fading in mobile systems (Hatzinakos and Llow, 1997)
    - Traffic modeling over comm. nets (Taqqu, 1996 – Petropulu, 2002)
  - Imaging Applications:
    - Modeling, compression, watermarking, classification, and image restoration in the DCT and Wavelet transform domains – The Cauchy detector, the KLD similarity function between SaS's for CBIR, the WIN-SAR processor, etc. (Tsakalides et al., 2000-2007)

SaS Applications in Astronomy/Physics

- Holtzmark, 1919: Distribution of the gravitational force created at a randomly chosen point by a given system of stars (considered to be a statistically homogeneous set of physical points which mutually interact according to the gravitation law). This distribution corresponds to a stable law with index $\alpha = 3/2$.

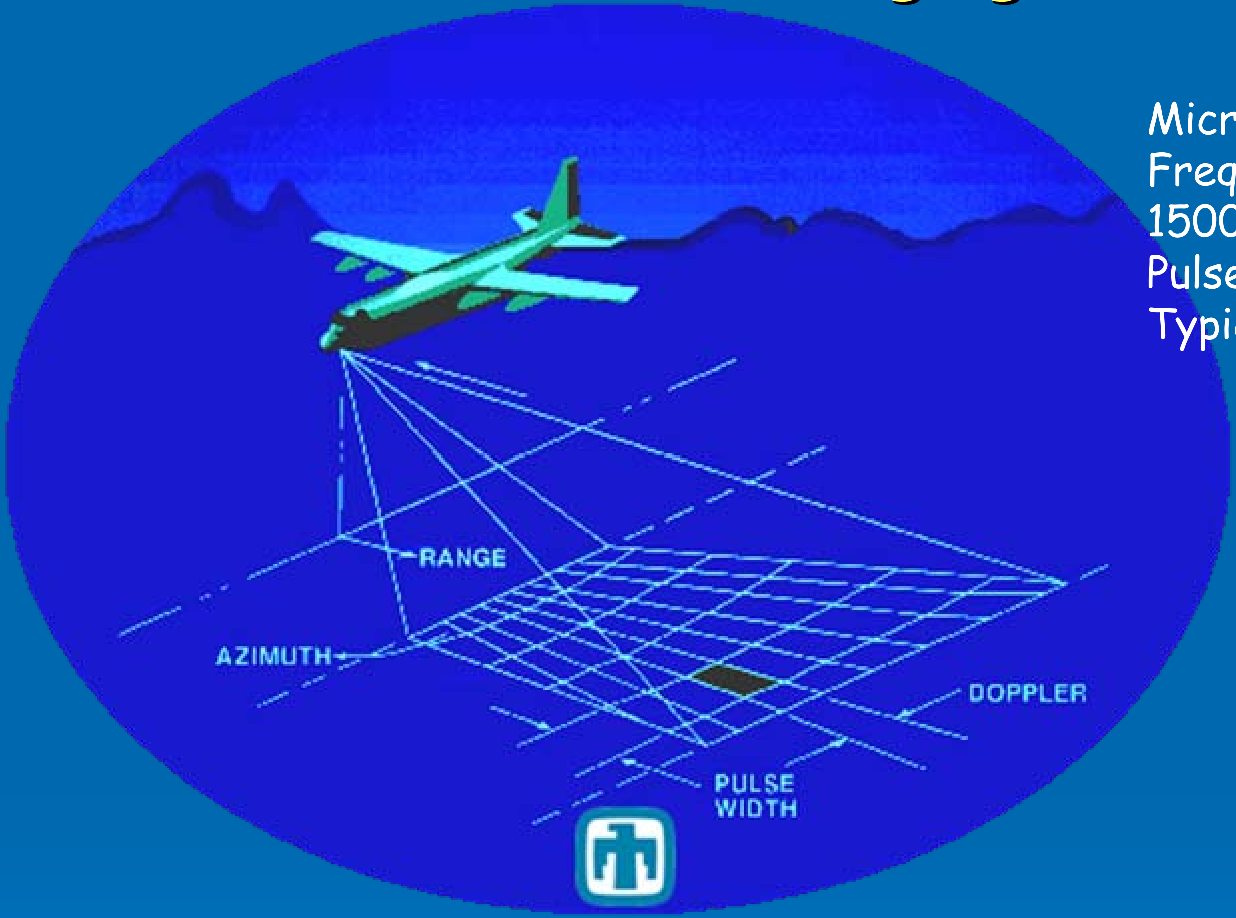
- Yanovsky, A.V. Chechkin, Schertzer, Tur (1999): "Fractional Fokker-Planck" equation, which includes fractional space differentiations, in order to encompass the wide class of anomalous diffusions due to a Levy stable stochastic forcing.

- Herranz, Kuruoglu, and L. Toffolatti, 2004: Distribution of unresolved point sources in Cosmic Microwave Background sky maps.

Multiscale methods for image processing:

The Wavelet-based Image-Denoising Nonlinear SAR (WIN-SAR) Processor

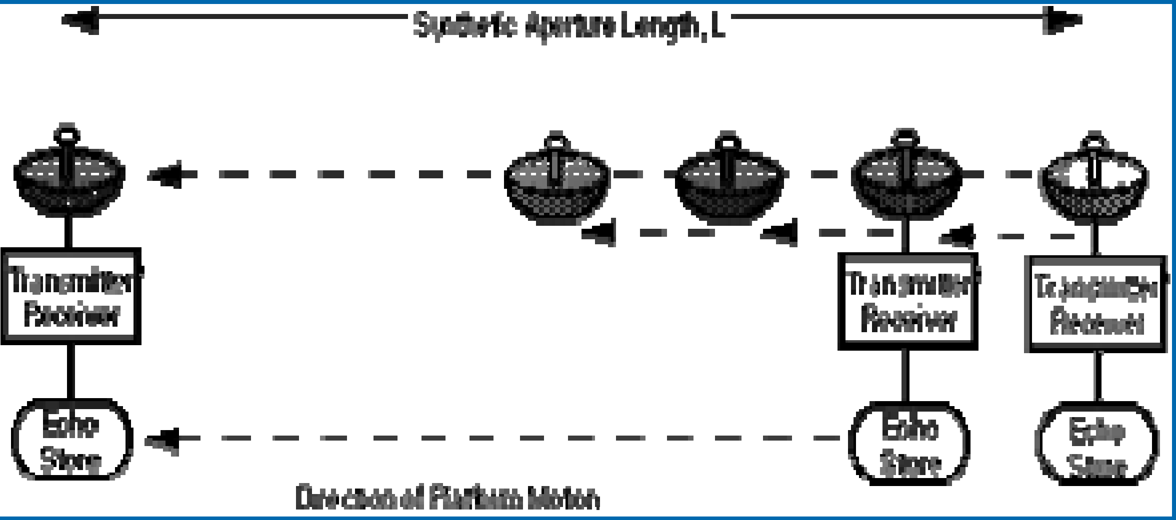
SAR Imaging Concept



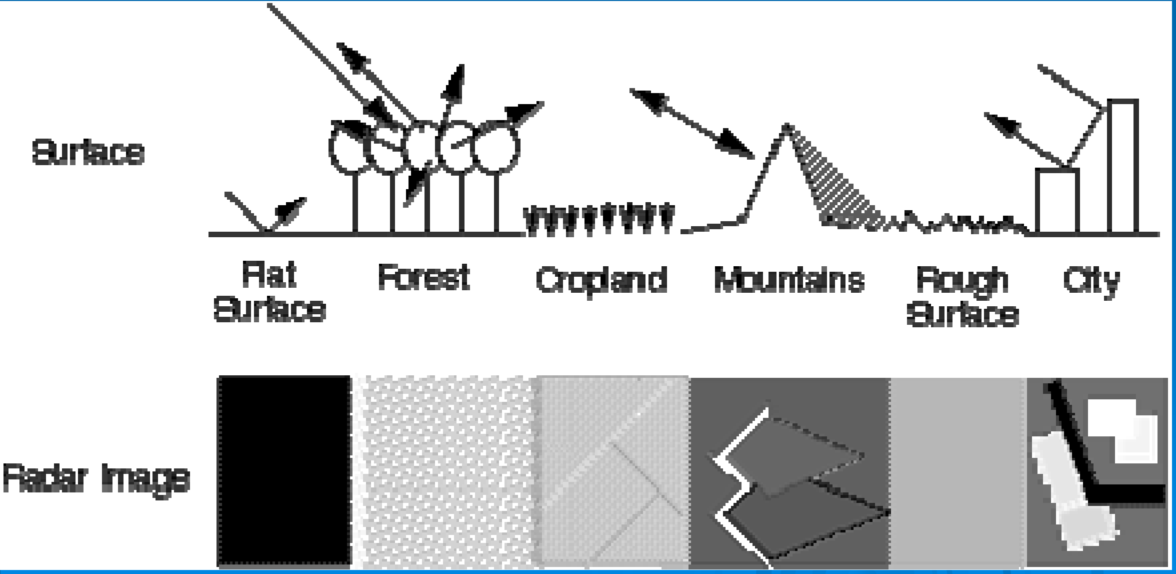
Microwave wavelenghts: 1 cm - 1 m
Frequency ranges: 300 MHz-30 GHz
1500 pulses/sec
Pulse duration: 10-50 µsecs
Typical Bandwidth: 10-200 MHz

SAR produces a two-dimensional (2-D) image. The *cross-track* dimension in the image is called *range* and is a measure of the "line-of-sight" distance from the radar to the target. The *along-track* dimension is called *azimuth* and is perpendicular to range.

SAR Imaging Concept



The length of the radar antenna determines the *resolution in the azimuth (along-track) direction* of the image: the longer the antenna, the finer the resolution in this dimension.



Each pixel in the image represents the radar *backscatter* for that area on the ground: objects approximately the size of the wavelength (or larger) appearing bright (i.e. rough) and objects smaller than the wavelength appearing dark (i.e. smooth)

SAR Qualities

- Synthetic aperture radar (SAR) systems take advantage of the long-range propagation characteristics of radar signals and the complex information processing capability of modern digital electronics to provide high resolution imagery.

- SAR complements photographic and other optical imaging capabilities because of the minimum constraints on time-of-day and atmospheric conditions and because of the unique responses of terrain and cultural targets to radar frequencies.

SIR-C/X-SAR Image of LA (space shuttle Endeavour, Oct. 1994)

- Area: 62x32 sq. miles
- L-band (24 cm) radar channel
- Horizontal polarizarion

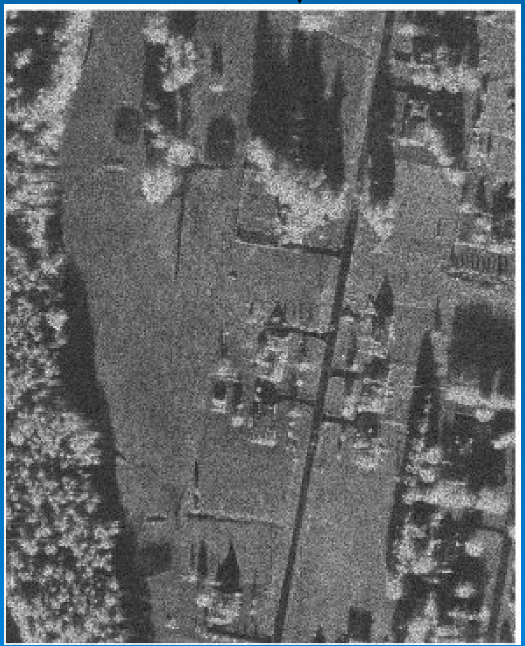
- Very dark grey: Pacific ocean, LAX, freeway system.
- Dark grey: mountain slops
- Lighter grey: suburban areas, low-density housing
- Bright white: high-rise buildings and housing alligned parallel to radar flight track

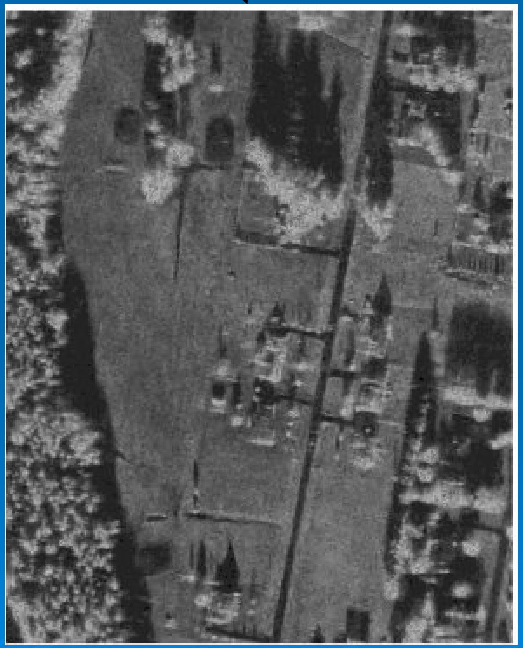
- Can be used to map fire scars in areas prone to brush fires, such as Los Angeles

Problem

SAR images are inherently affected by multiplicative speckle noise, due to the coherent nature of the scattering phenomenon as well as additive noise:

$$I(x,y) = S(x,y) \cdot n_m(x,y) + n_a(x,y)$$

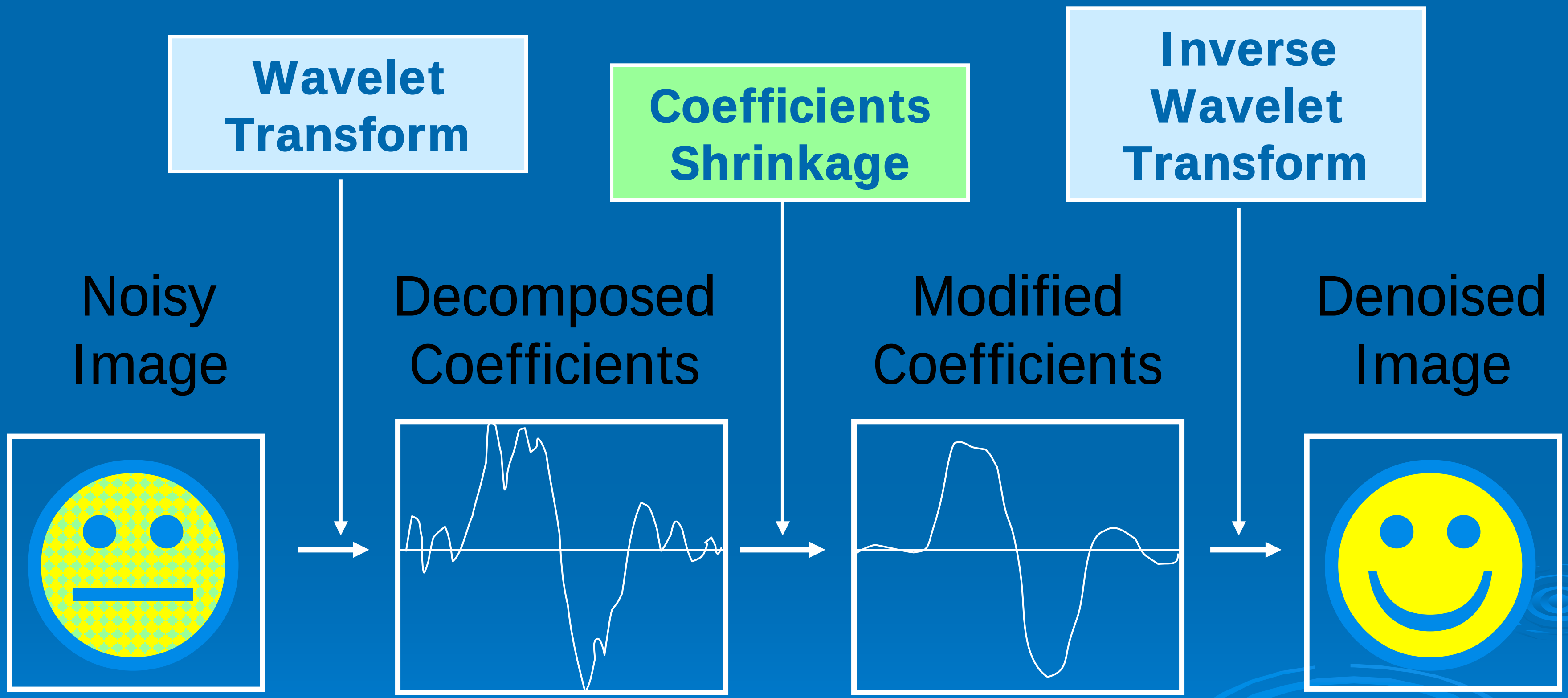




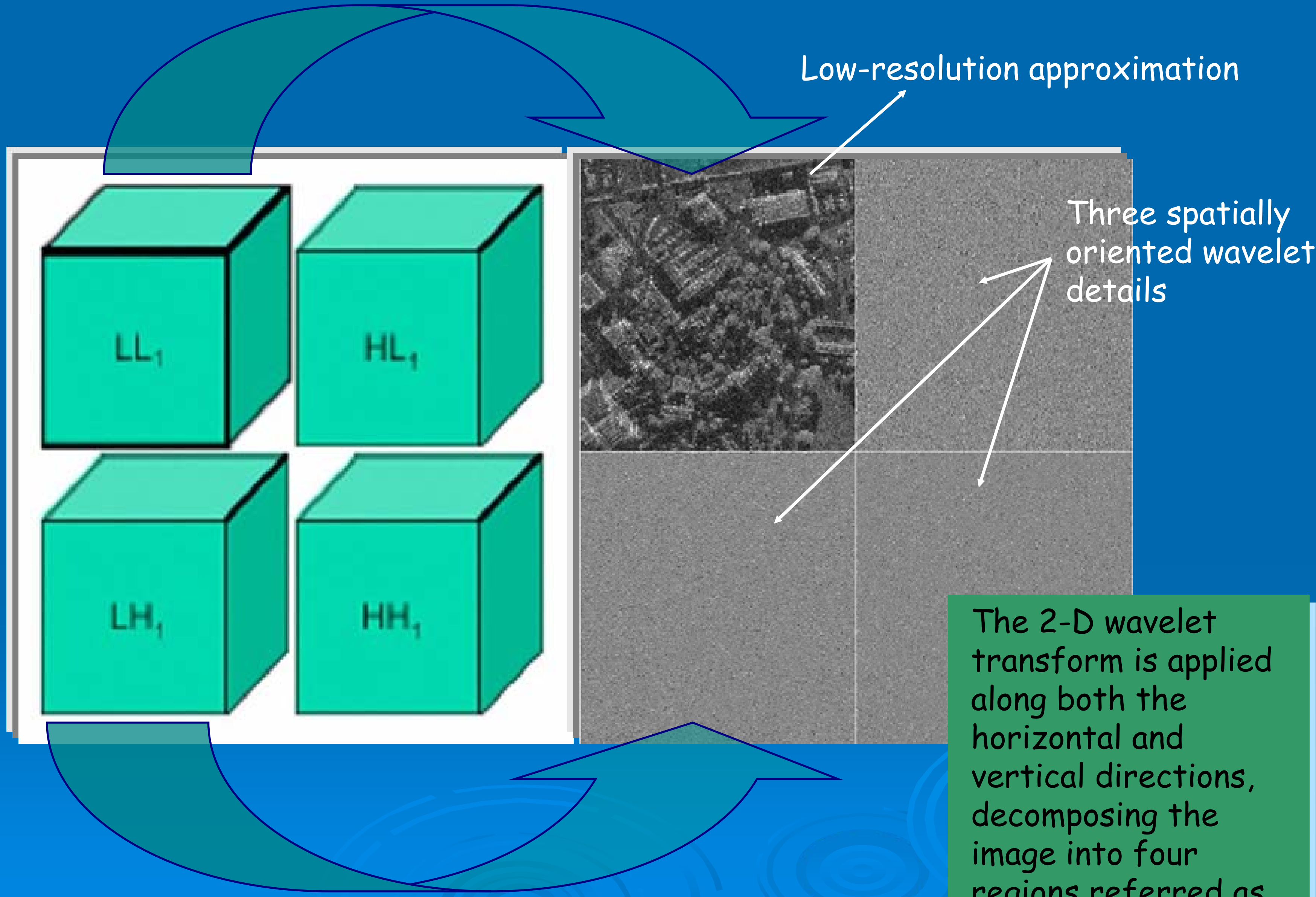
Speckle Noise (multiplicative): unit-mean, log-normal distributed.

Need to balance between speckle suppression and signal detail preservation!!!

Wavelets for Image Denoising

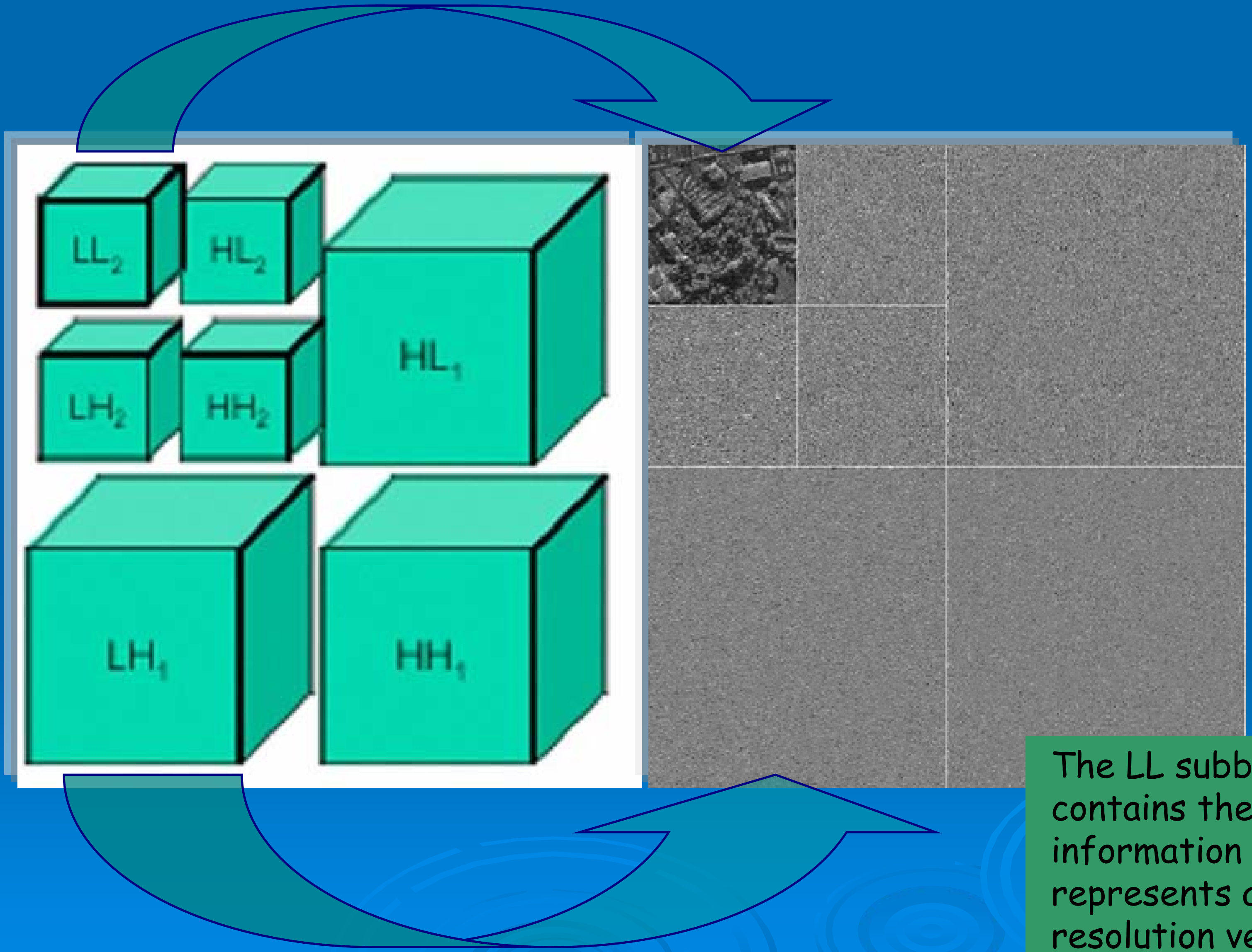


Multiresolution decomposition – 1st level



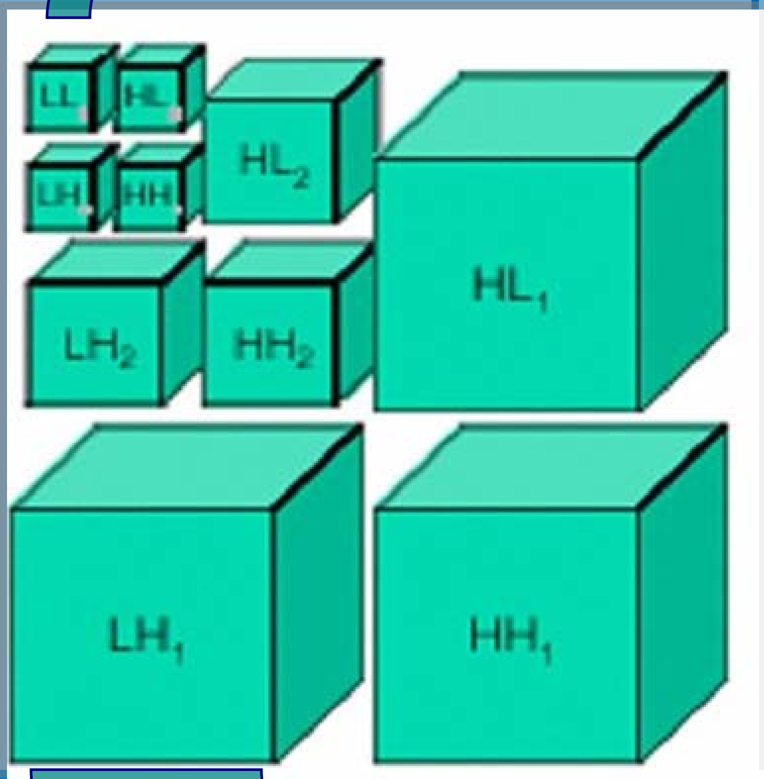
The 2-D wavelet transform is applied along both the horizontal and vertical directions, decomposing the image into four regions referred as image subbands.

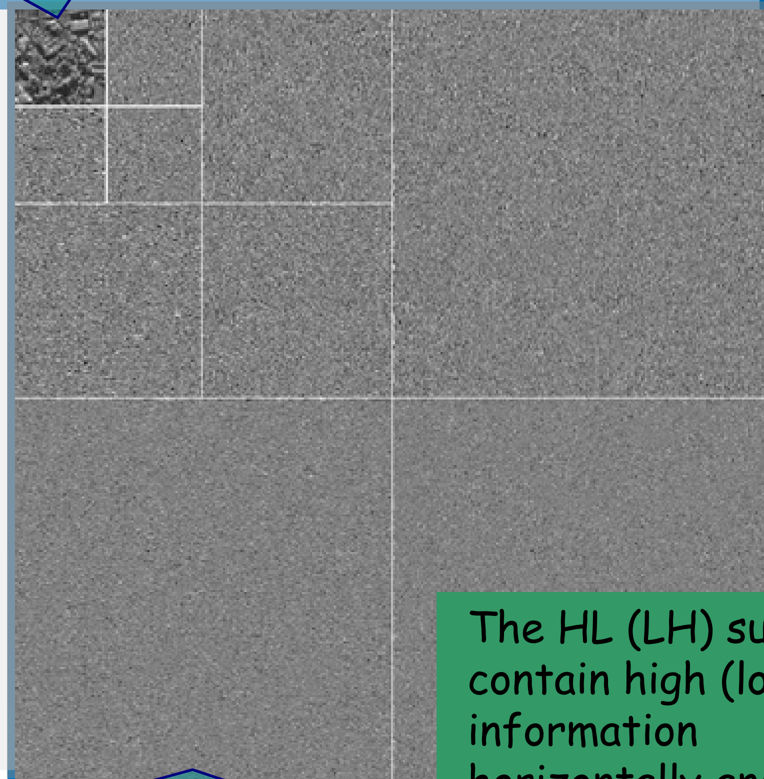
Multiresolution decomposition – 2nd level



The LL subband contains the low-pass information and it represents a low resolution version of the original image.

Multiresolution decomposition – 3rd level





The HL (LH) subbands contain high (low) pass information horizontally and low (high) pass information vertically. The HH subbands contain high-pass information in both directions.

Previous Work in Wavelet-based Image Denoising

- Donoho's pioneering work: "Denoising by soft-thresholding" IEEE Trans. Inf. Theory 1995
- Simoncelli's "Noise removal via Bayesian wavelet coding," 1996
- Gagnon & Jouan's wavelet coefficient shrinkage (WCS) filter, 1997
- Simoncelli's work on texture synthesis, 1999
- Sadler's multiscale point-wise product technique, 1999
- Achim's work on heavy-tailed modeling, 2001
- Pizurica's work on inter & intra-scale statistical modeling, 2002

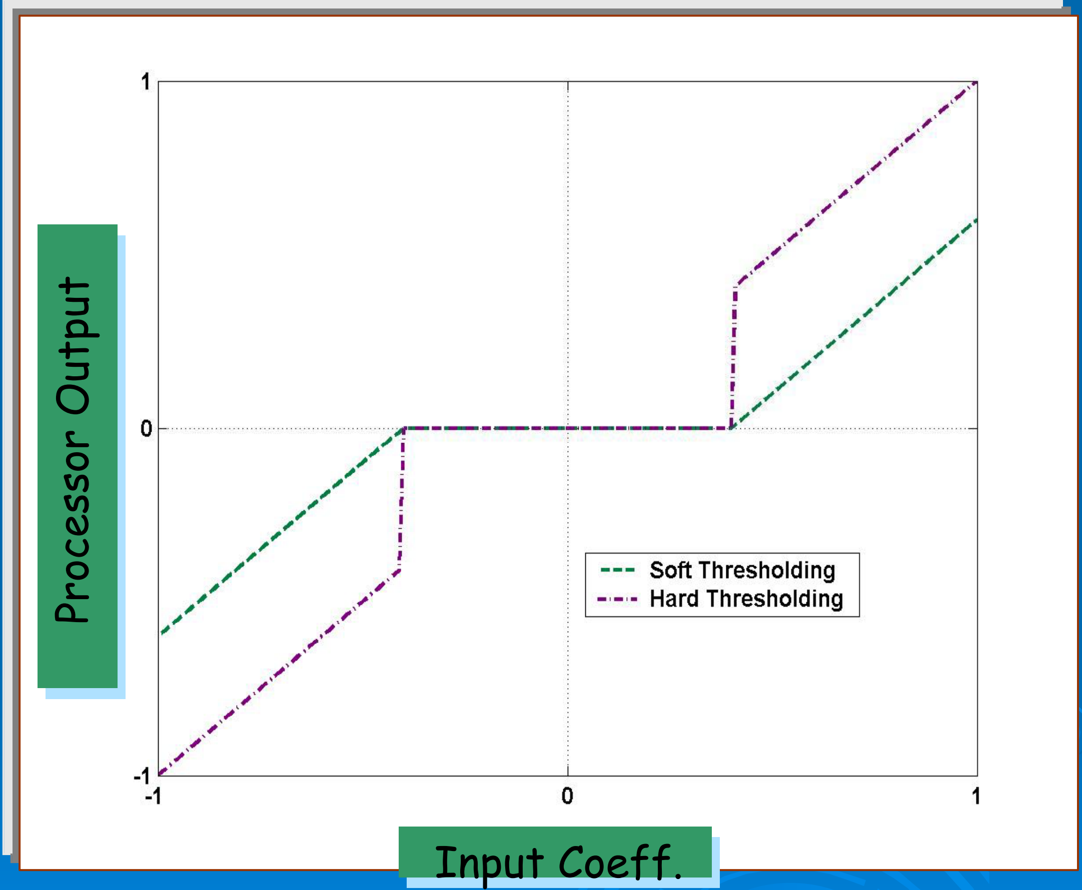
Wavelet Shrinkage Methods

- Soft Thresholding

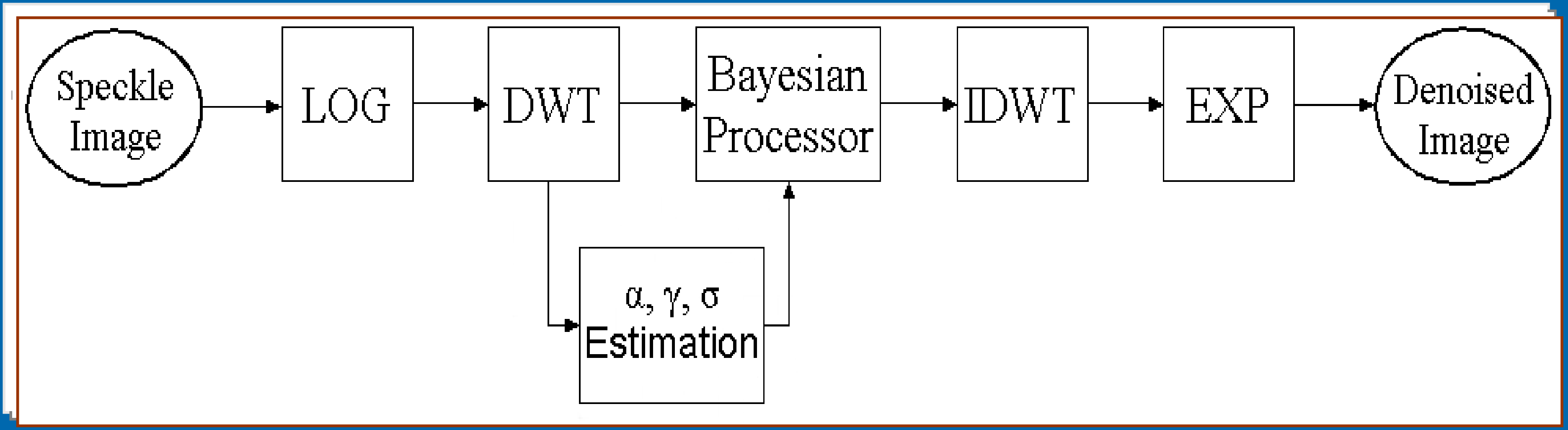
$$T_s^{soft}(s) = \begin{cases} \operatorname{sgn}(s)(|s| - t), & |s| > t \\ 0, & |s| \le t \end{cases}$$

- Hard Thresholding

$$T_s^{hard}(s) = \begin{cases} s, & |s| > t \\ 0, & |s| \le t \end{cases}$$



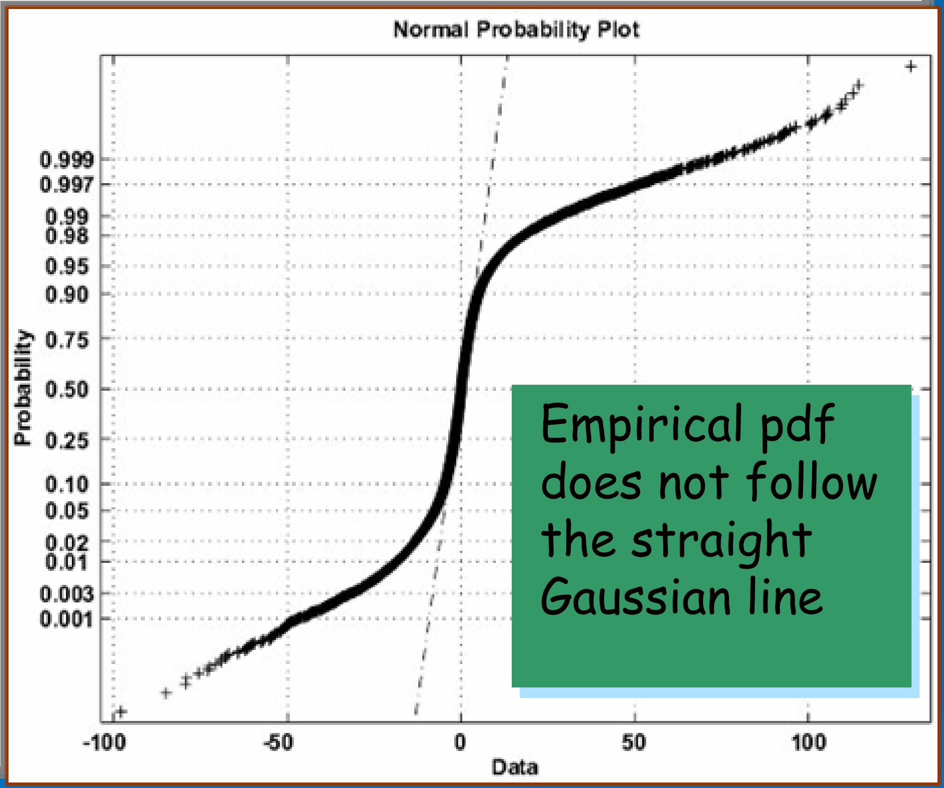
The WIN-SAR Processor

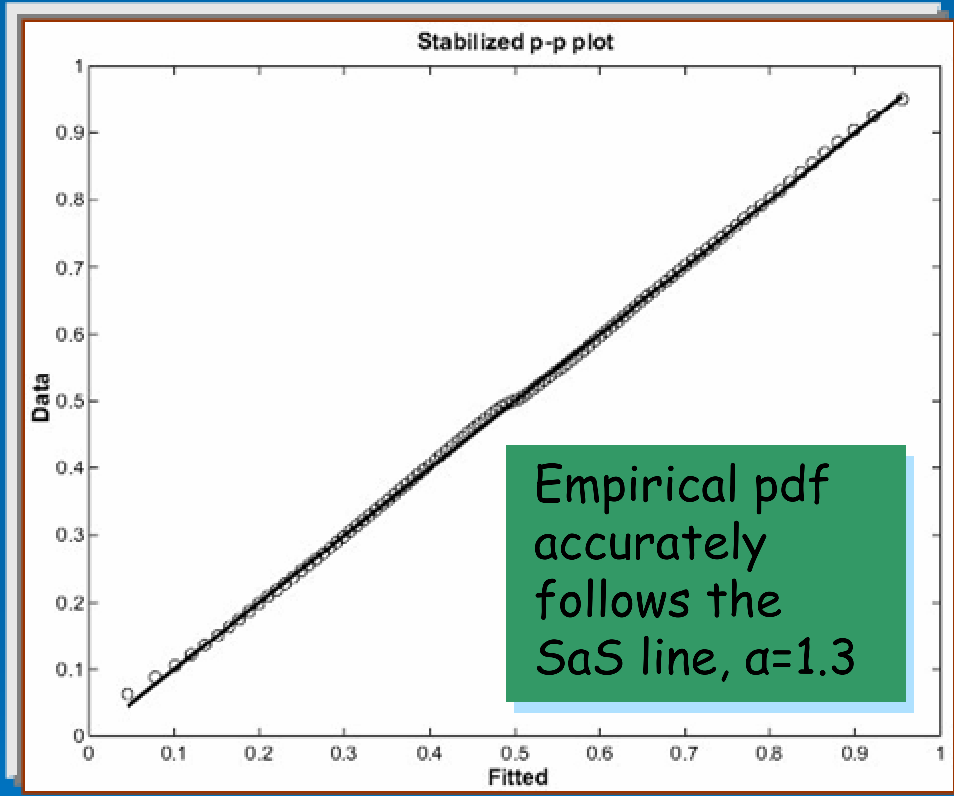


WIN-SAR fundamentals:

1. Wavelet transform the speckle SAR image.
2. SaS modeling of signal wavelet coefficients.
3. Bayesian processing of the coefficients in every level of decomposition.

Wavelet Coefficients Modeling (1)

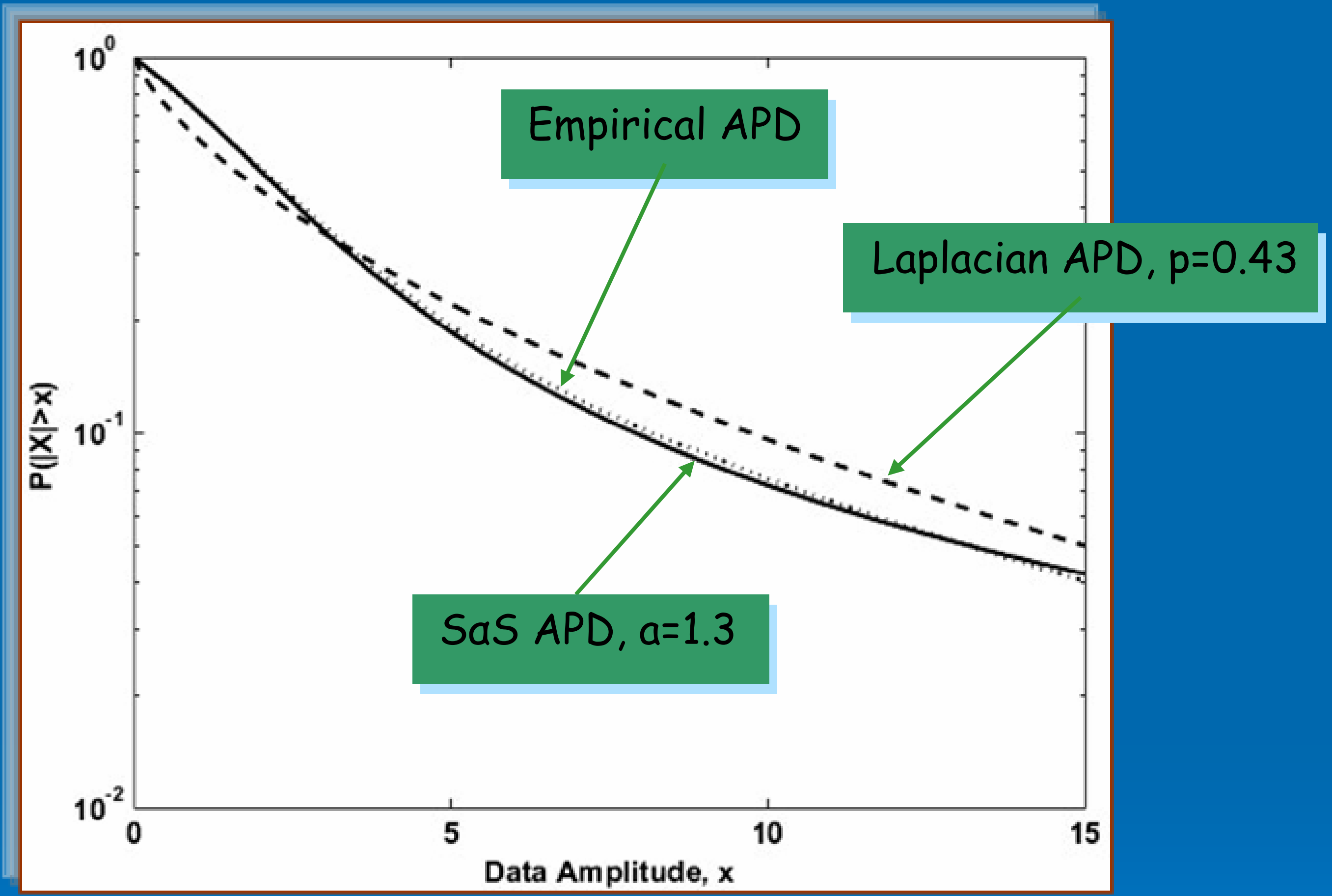




Normal and SaS probability plots of the vertical subband at the first level of decomposition of the image HB06158 from the MSTAR* collection.

* http://www.mvlab.wpafb.af.mil/public/sdms/

Wavelet Coefficients Modeling (2)



Amplitude Probability Density (APD) plot for the data of the previous slide: The SaS provides an excellent fit to both the **mode** and the **tails** of the empirical distribution.

SaS Modeling of Wavelet Subband Coefficients

Level	Image Subbands		
	Horizontal	Vertical	Diagonal
1	1.239	1.283	1.302
2	1.418	1.125	1.295
3	1.286	1.019	1.380

The tabulated key parameter α defines the degree of non-Gaussianity as deviations from the value $\alpha = 2$.

The WIN-SAR MAE Bayesian Estimator

- After applying the DWT: $d_{j,k}^{i} = s_{j,k}^{i} + \xi_{j,k}^{i}$

- The Bayes risk estimator of s minimizes the conditional risk, i.e., the loss function averaged over the conditional distribution of s given the measured wavelet coeffs:

$$\hat{s}(d) = \underset{\hat{s}(d)}{\operatorname{argmin}} \int \left| s - \hat{s}(d) \right| \cdot P_{s|d}(s \mid d) \cdot ds$$

- The <u>mean absolute error</u> (MAE) estimator is the conditional median of s, given d, which coincides with the conditional mean (due to the symmetry of the distributions):

$$\hat{s}(d) = \int s \cdot P_{s|d}(s \mid d) \cdot ds = \frac{\int P_{d/s}(d/s) P(s) s \cdot ds}{\int P_{d/s}(d/s) P(s) \cdot ds}$$

The WIN-SAR MAE Bayesian Estimator

$$\hat{s}(d) = \frac{\int P_\xi(d-s)P(s)s\cdot ds}{\int P_\xi(d-s)P(s)\cdot ds} = \frac{\int P_\xi(\xi)P(s)s\cdot ds}{\int P_\xi(\xi)P(s)\cdot ds}$$

- Signal Parameter Estimation - by means of a LS fitting in the characteristic function domain:

$$\{\hat{a}_s, \hat{\gamma}_s, \hat{\sigma}\} = \underbrace{\arg\min}_{\hat{a}_s, \hat{\gamma}_s, \hat{\sigma}} \sum_i^n \left[\Phi_d(\omega_i) - \Phi_{d\,e}(\omega_i)\right]^2$$

where:

$$\Phi_d(\omega) = \exp(-\gamma_s |\omega|^{\alpha_s}) \cdot \exp(-\frac{\sigma^2}{2}|\omega|^2)$$

WIN-SAR MAE Processor I/O Curves

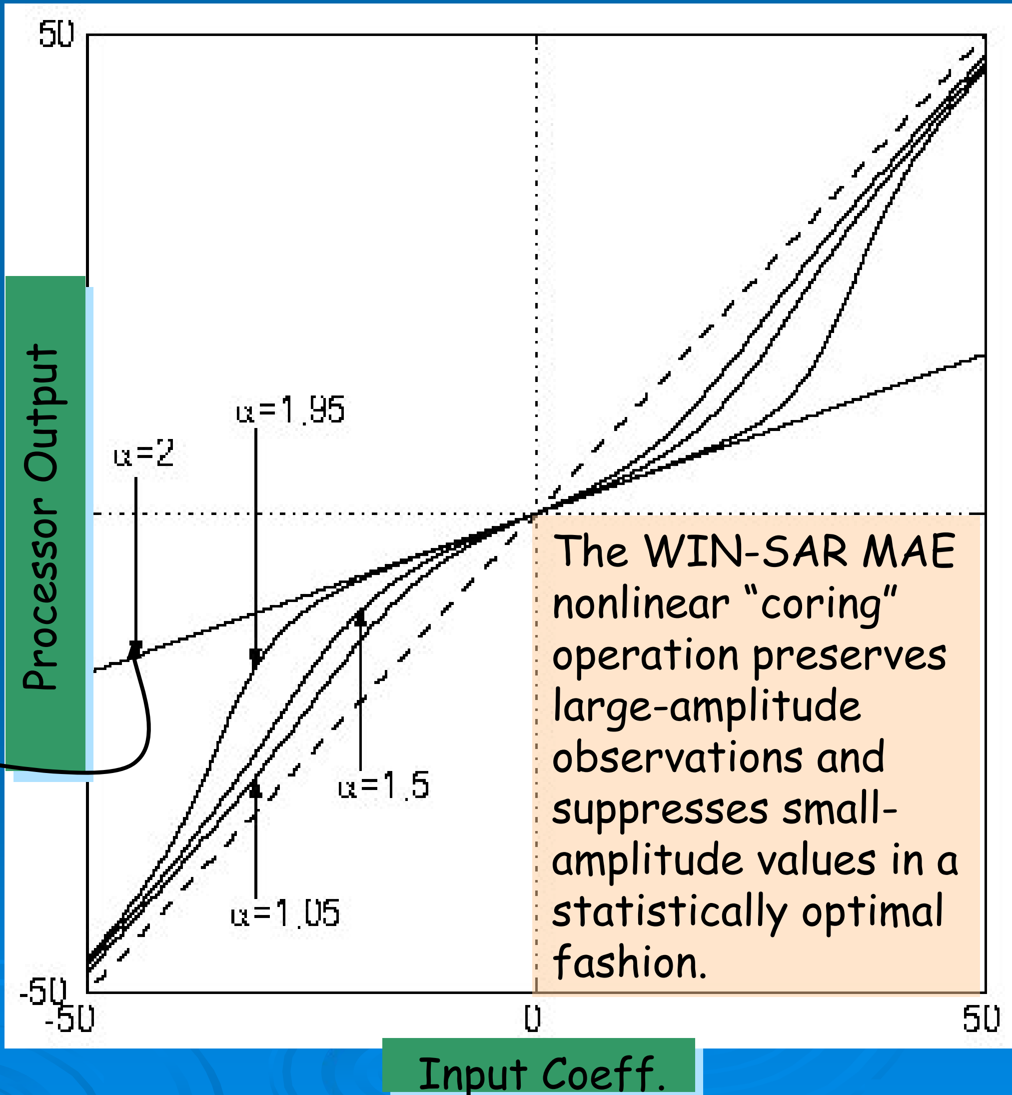
Bayesian Processing:

$$\hat{s}(d) = \frac{\int P_{\xi}(\xi) P(s) s \cdot ds}{\int P_{\xi}(\xi) P(s) \cdot ds}$$

Only for a=2 (Gaussian signal), the processing is a simple linear rescaling of the measurement:

$$\hat{s}(d) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2} d$$

For a given ratio γ/σ, the amount of shrinkage decreases as α decreases: The smaller the value of α, the heavier the tails of the signal PDF and the greater the probability that the measured value is due to the signal.



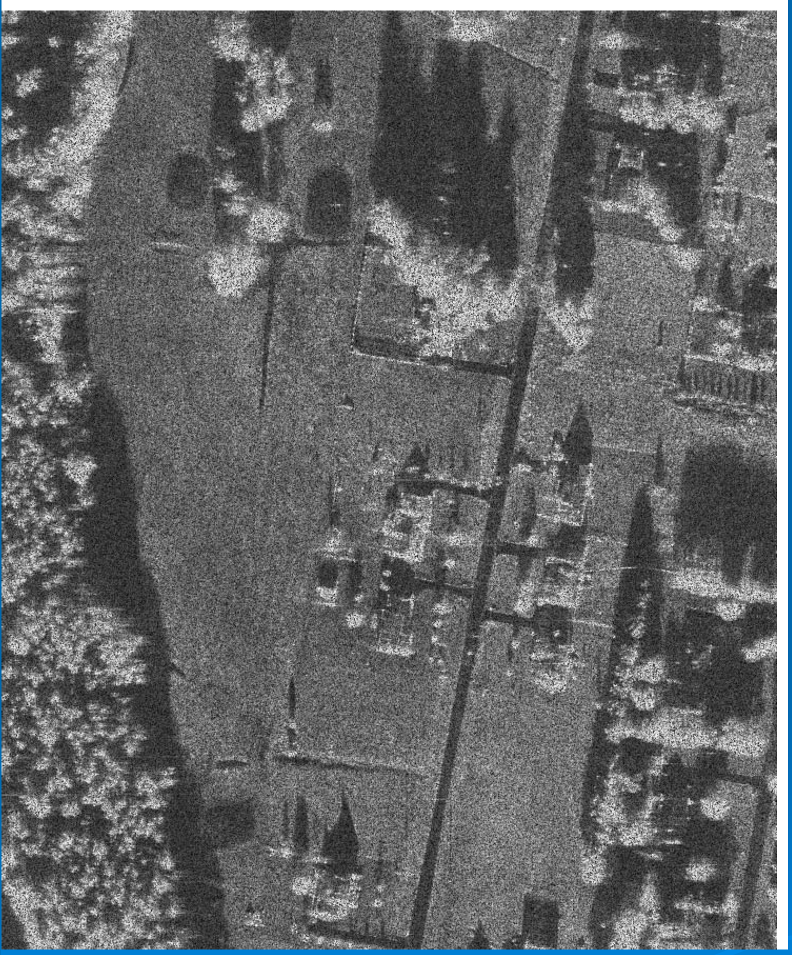
Real SAR Imagery Results (2)

Urban scene (dense set of large cross-section targets w. intermingled tree shadows

WIN-SAR

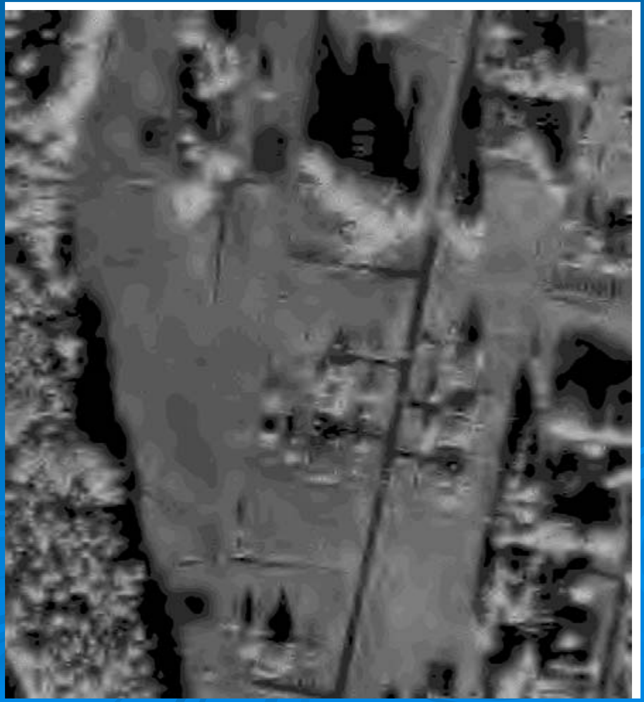
Shoft-Thresholding

Real SAR Imagery Results (3)



Rural scene

WIN-SAR



Shoft-Thresholding

Conclusions

1. Motivated the use of alpha-stable processes as appropriate modeling tools for a large class of applications.
2. Designed and tested Bayesian processors and found them more effective than traditional wavelet shrinkage methods, both in terms of speckle reduction and signal detail preservation.
3. Proposed processors based on solid statistical theory: do not depend on the use of any *ad hoc* thresholding parameter.
4. Can the alpha-stable model and the associated detection, estimation and classification techniques be of use in Astronomical Data Analysis???